



wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 01:06 PM GMT

PDB ID : 4HG4
Title : Crystal structure of Fab 2G1 in complex with a H2N2 influenza virus hemagglutinin
Authors : Xu, R.; Wilson, I.A.
Deposited on : 2012-10-06
Resolution : 3.20 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

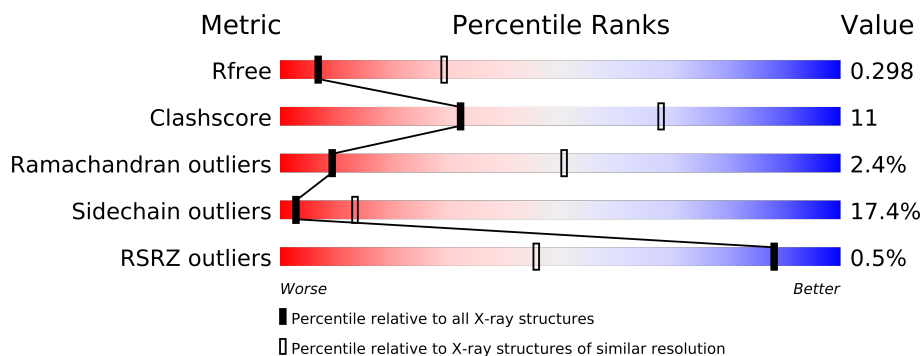
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

1 Overall quality at a glance

The reported resolution of this entry is 3.20 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	66092	1824 (3.30-3.10)
Clashscore	79885	1078 (3.26-3.14)
Ramachandran outliers	78287	1059 (3.26-3.14)
Sidechain outliers	78261	1058 (3.26-3.14)
RSRZ outliers	66119	1825 (3.30-3.10)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	327	
1	B	327	
1	C	327	
1	D	327	
1	E	327	
1	F	327	
1	G	327	
1	H	327	
1	I	327	
2	a	174	
2	b	174	
2	c	174	
2	d	174	
2	e	174	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	f	174	
2	g	174	
2	h	174	
2	i	174	
3	J	223	
3	L	223	
3	N	223	
3	P	223	
3	R	223	
3	T	223	
3	V	223	
3	X	223	
3	Z	223	
4	K	214	
4	M	214	
4	O	214	
4	Q	214	
4	S	214	
4	U	214	
4	W	214	
4	Y	214	
4	z	214	

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 58742 atoms, of which 0 are hydrogen and 0 are deuterium.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Hemagglutinin HA1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	323	Total	C	N	O	S	0	0	0
			2531	1593	438	485	15			
1	B	319	Total	C	N	O	S	0	0	0
			2503	1577	433	478	15			
1	C	324	Total	C	N	O	S	0	0	0
			2539	1599	439	486	15			
1	D	324	Total	C	N	O	S	0	0	0
			2539	1599	439	486	15			
1	E	323	Total	C	N	O	S	0	0	0
			2531	1593	438	485	15			
1	F	324	Total	C	N	O	S	0	0	0
			2539	1599	439	486	15			
1	G	324	Total	C	N	O	S	0	0	0
			2539	1599	439	486	15			
1	H	325	Total	C	N	O	S	0	0	0
			2548	1604	440	489	15			
1	I	324	Total	C	N	O	S	0	0	0
			2539	1599	439	486	15			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	9	PRO	-	EXPRESSION TAG	UNP Q67085
B	9	PRO	-	EXPRESSION TAG	UNP Q67085
C	9	PRO	-	EXPRESSION TAG	UNP Q67085
D	9	PRO	-	EXPRESSION TAG	UNP Q67085
E	9	PRO	-	EXPRESSION TAG	UNP Q67085
F	9	PRO	-	EXPRESSION TAG	UNP Q67085
G	9	PRO	-	EXPRESSION TAG	UNP Q67085
H	9	PRO	-	EXPRESSION TAG	UNP Q67085
I	9	PRO	-	EXPRESSION TAG	UNP Q67085

- Molecule 2 is a protein called Hemagglutinin HA2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	a	171	Total	C	N	O	S	0	0	0
			1387	866	236	276	9			
2	b	167	Total	C	N	O	S	0	0	0
			1345	836	229	271	9			
2	c	172	Total	C	N	O	S	0	0	0
			1396	871	237	279	9			
2	d	172	Total	C	N	O	S	0	0	0
			1396	871	237	279	9			
2	e	172	Total	C	N	O	S	0	0	0
			1396	871	237	279	9			
2	f	172	Total	C	N	O	S	0	0	0
			1396	871	237	279	9			
2	g	172	Total	C	N	O	S	0	0	0
			1396	871	237	279	9			
2	h	167	Total	C	N	O	S	0	0	0
			1355	845	231	270	9			
2	i	172	Total	C	N	O	S	0	0	0
			1396	871	237	279	9			

- Molecule 3 is a protein called Fab 2G1 heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	J	115	Total	C	N	O	S	0	0	0
			871	555	142	171	3			
3	L	212	Total	C	N	O	S	0	0	0
			1578	1004	258	311	5			
3	N	223	Total	C	N	O	S	0	0	0
			1648	1043	271	328	6			
3	P	161	Total	C	N	O	S	0	0	0
			1207	770	195	239	3			
3	R	214	Total	C	N	O	S	0	0	0
			1589	1010	260	314	5			
3	T	213	Total	C	N	O	S	0	0	0
			1580	1005	259	311	5			
3	V	117	Total	C	N	O	S	0	0	0
			883	564	144	172	3			
3	X	164	Total	C	N	O	S	0	0	0
			1232	785	201	242	4			
3	Z	167	Total	C	N	O	S	0	0	0
			1262	807	205	246	4			

- Molecule 4 is a protein called Fab 2G1 light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	K	106	Total	C	N	O	S	0	0	0
			826	518	140	166	2			
4	M	214	Total	C	N	O	S	0	0	0
			1661	1037	282	337	5			
4	O	214	Total	C	N	O	S	0	0	0
			1661	1037	282	337	5			
4	Q	157	Total	C	N	O	S	0	0	0
			1217	760	206	248	3			
4	S	211	Total	C	N	O	S	0	0	0
			1642	1027	279	332	4			
4	U	211	Total	C	N	O	S	0	0	0
			1642	1027	279	332	4			
4	W	105	Total	C	N	O	S	0	0	0
			818	512	139	165	2			
4	Y	109	Total	C	N	O	S	0	0	0
			853	534	147	170	2			
4	z	119	Total	C	N	O	S	0	0	0
			915	568	155	190	2			

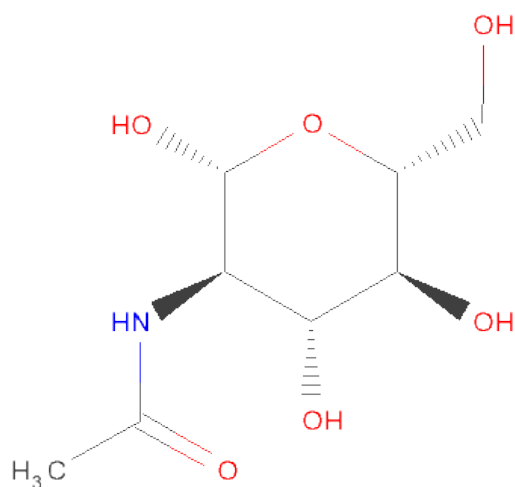
- Molecule 5 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	2	Total	C	N	O	0	0
			28	16	2	10		
5	B	2	Total	C	N	O	0	0
			28	16	2	10		
5	D	2	Total	C	N	O	0	0
			28	16	2	10		
5	E	2	Total	C	N	O	0	0
			28	16	2	10		
5	F	2	Total	C	N	O	0	0
			28	16	2	10		
5	H	2	Total	C	N	O	0	0
			28	16	2	10		
5	H	2	Total	C	N	O	0	0
			28	16	2	10		
5	I	2	Total	C	N	O	0	0
			28	16	2	10		

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	9	PRO	-	EXPRESSION TAG	UNP Q67085
B	9	PRO	-	EXPRESSION TAG	UNP Q67085
D	9	PRO	-	EXPRESSION TAG	UNP Q67085
E	9	PRO	-	EXPRESSION TAG	UNP Q67085
F	9	PRO	-	EXPRESSION TAG	UNP Q67085
H	9	PRO	-	EXPRESSION TAG	UNP Q67085
H	9	PRO	-	EXPRESSION TAG	UNP Q67085
H	9	PRO	-	EXPRESSION TAG	UNP Q67085
I	9	PRO	-	EXPRESSION TAG	UNP Q67085

- Molecule 6 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	A	1	Total	C	N	O	0	0
			14	8	1	5		
6	C	1	Total	C	N	O	0	0
			14	8	1	5		
6	E	1	Total	C	N	O	0	0
			14	8	1	5		
6	F	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 7 is a polymer of unknown type called SUGAR (3-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
7	C	3	Total	C	N	O	0	0
			39	22	2	15		
7	G	3	Total	C	N	O	0	0
			39	22	2	15		

There are 2 discrepancies between the modelled and reference sequences:

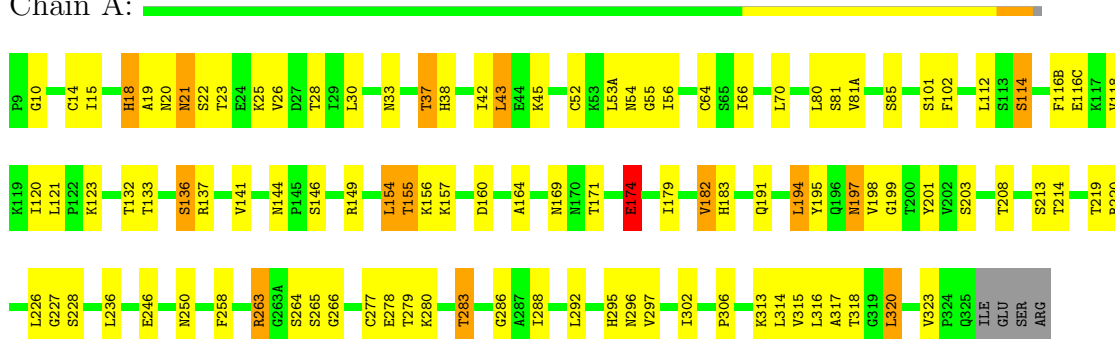
Chain	Residue	Modelled	Actual	Comment	Reference
C	9	PRO	-	EXPRESSION TAG	UNP Q67085
G	9	PRO	-	EXPRESSION TAG	UNP Q67085

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

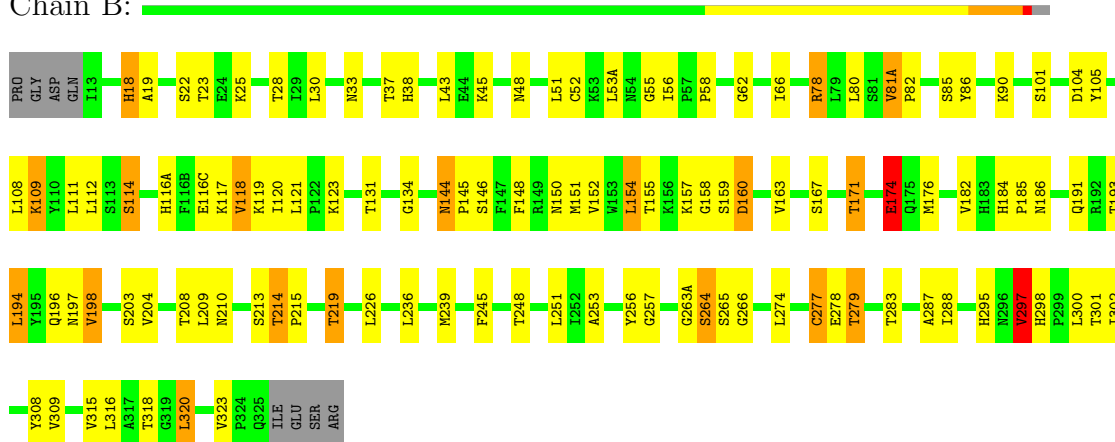
• Molecule 1: Hemagglutinin HA1

Chain A:



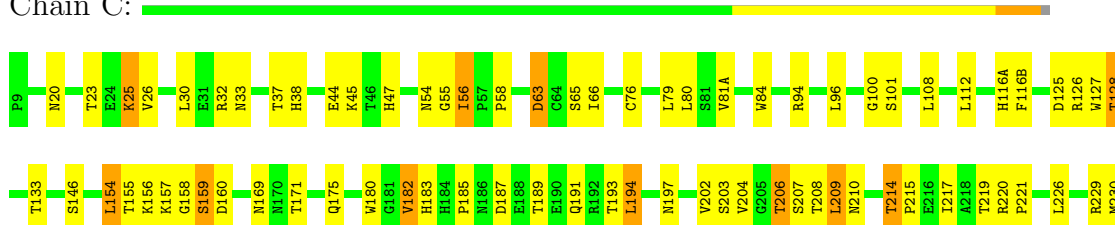
• Molecule 1: Hemagglutinin HA1

Chain B:



• Molecule 1: Hemagglutinin HA1

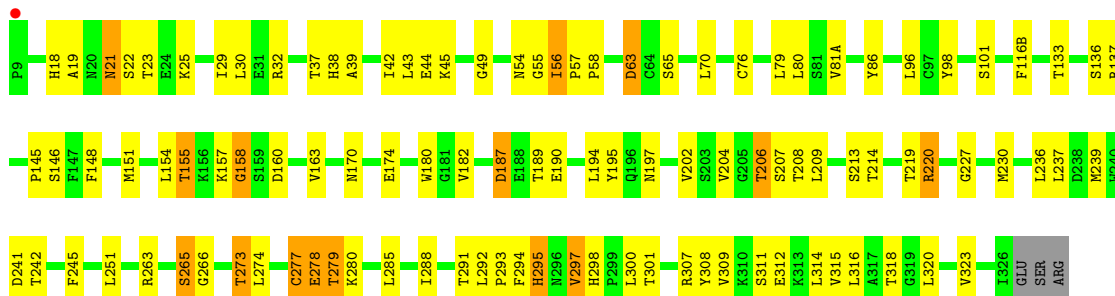
Chain C:





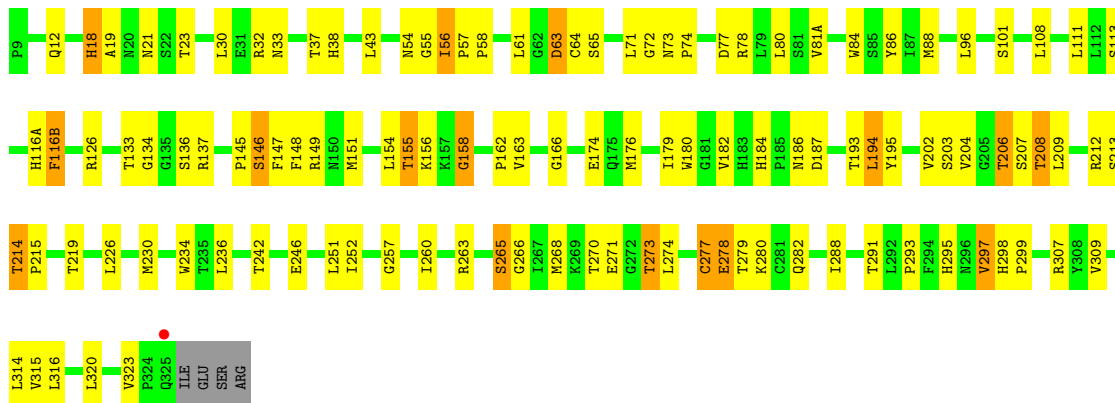
• Molecule 1: Hemagglutinin HA1

Chain D:



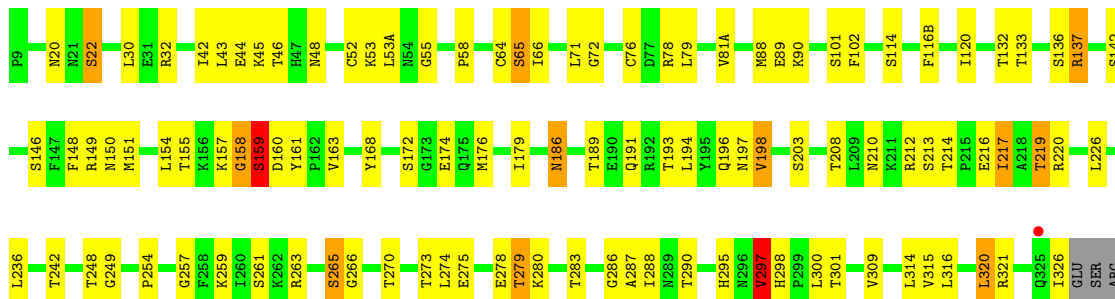
• Molecule 1: Hemagglutinin HA1

Chain E:



• Molecule 1: Hemagglutinin HA1

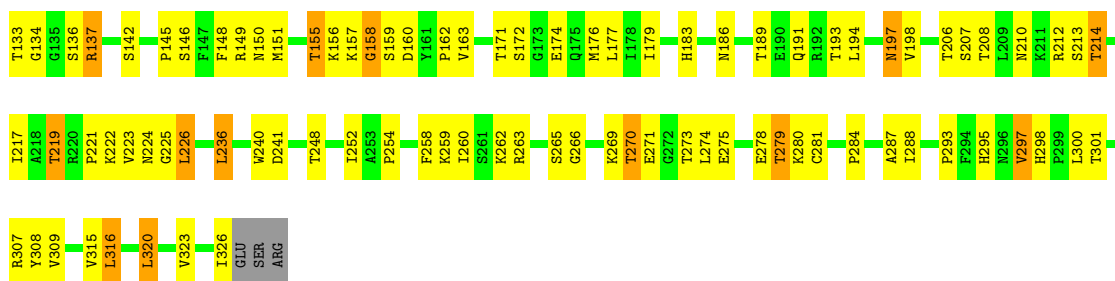
Chain F:



• Molecule 1: Hemagglutinin HA1

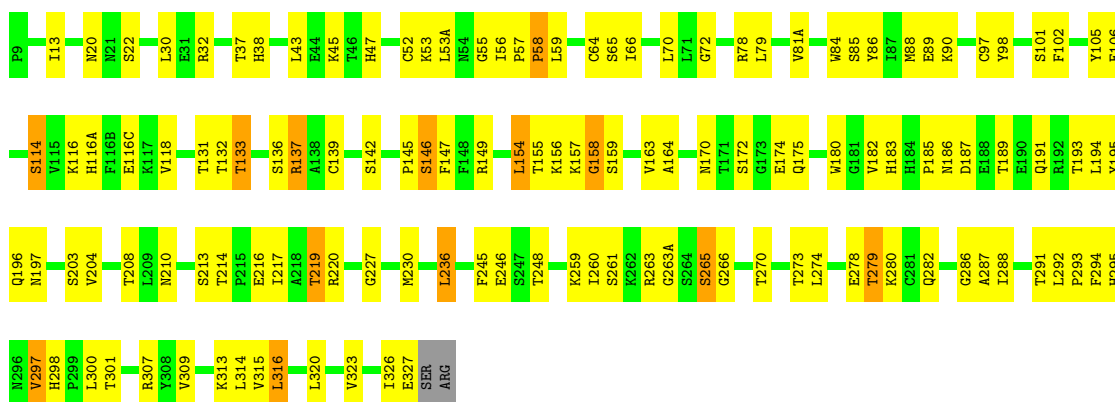
Chain G:





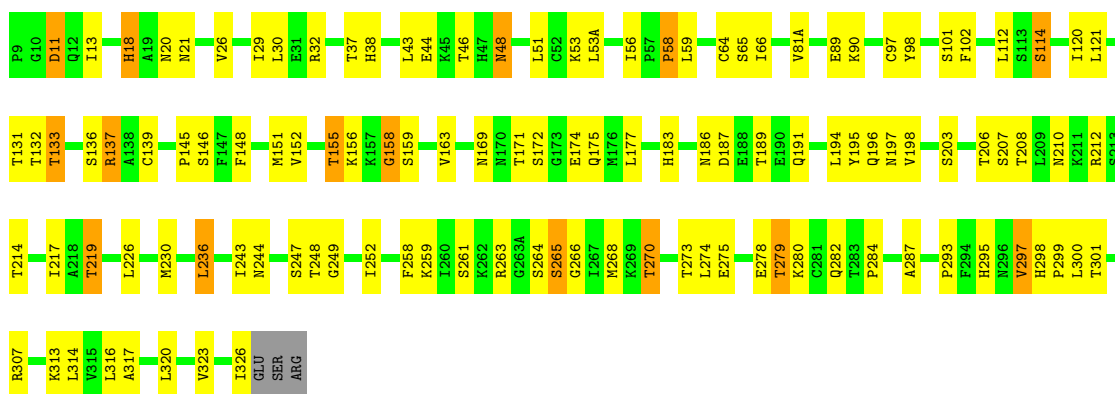
• Molecule 1: Hemagglutinin HA1

Chain H:



• Molecule 1: Hemagglutinin HA1

Chain I:



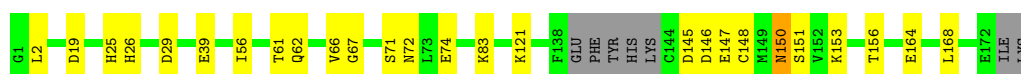
• Molecule 2: Hemagglutinin HA2

Chain a:



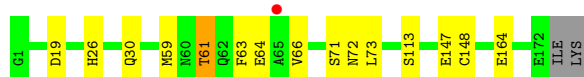
• Molecule 2: Hemagglutinin HA2

Chain b:



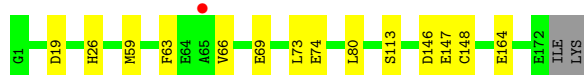
- Molecule 2: Hemagglutinin HA2

Chain c:



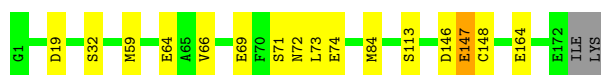
- Molecule 2: Hemagglutinin HA2

Chain d:



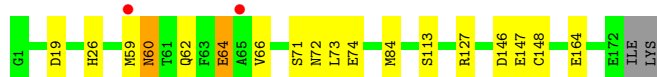
- Molecule 2: Hemagglutinin HA2

Chain e:



- Molecule 2: Hemagglutinin HA2

Chain f:



- Molecule 2: Hemagglutinin HA2

Chain g:



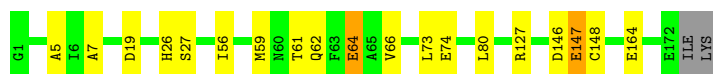
- Molecule 2: Hemagglutinin HA2

Chain h:



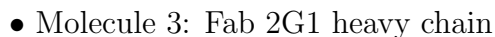
- Molecule 2: Hemagglutinin HA2

Chain i:

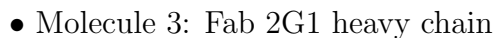
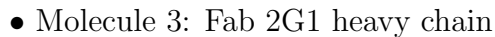
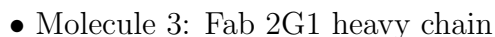


- Molecule 3: Fab 2G1 heavy chain

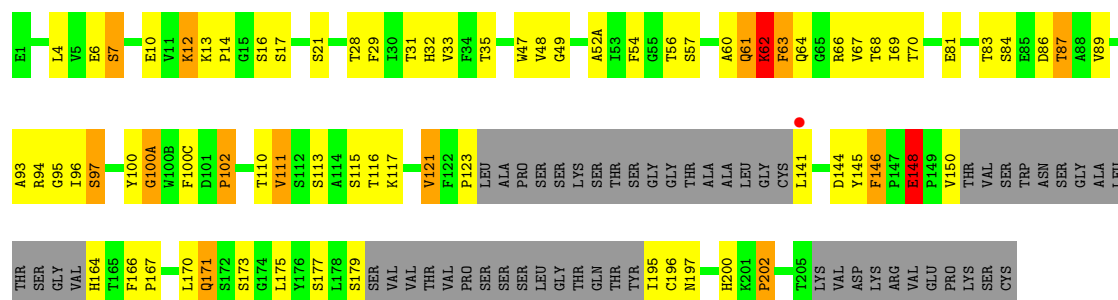
Chain J:



Age Group	Percentage of Respondents
18-29	95%
30-49	87%
50-64	77%
65+	63%

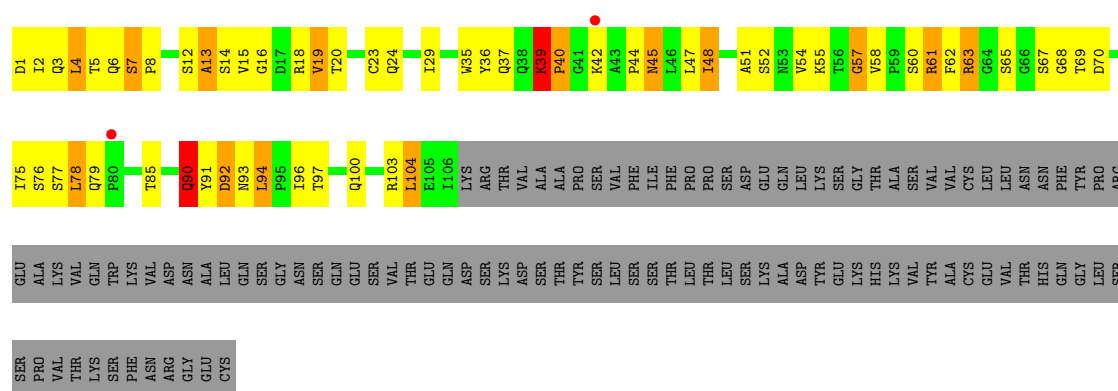


Chain Z:



- Molecule 4: Fab 2G1 light chain

Chain K:



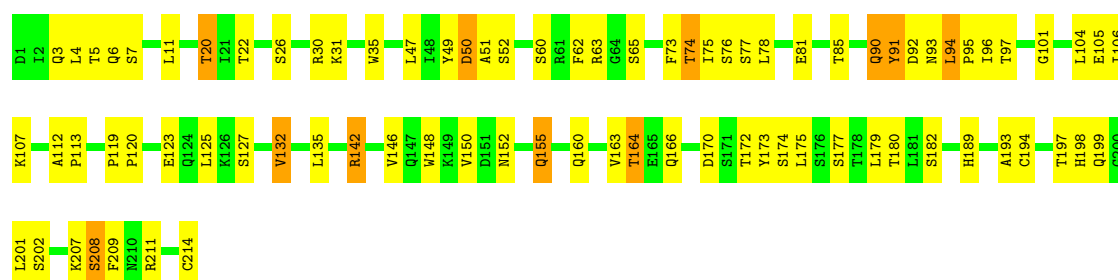
- Molecule 4: Fab 2G1 light chain

Chain M:



- Molecule 4: Fab 2G1 light chain

Chain O:





Chain Y:



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	126.84Å 133.14Å 812.99Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.13 – 3.20 50.01 – 3.16	Depositor EDS
% Data completeness (in resolution range)	90.2 (48.13-3.20) 87.4 (50.01-3.16)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	6.22 (at 3.19Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.2_869)	Depositor
R, R_{free}	0.249 , 0.302 0.246 , 0.298	Depositor DCC
R_{free} test set	10328 reflections (5.28%)	DCC
Wilson B-factor (Å ²)	55.3	Xtriage
Anisotropy	0.019	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 16.5	EDS
Estimated twinning fraction	0.012 for k,h,-l	Xtriage
L-test for twinning	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 206252 reflections	Xtriage
F_o, F_c correlation	0.86	EDS
Total number of atoms	58742	wwPDB-VP
Average B, all atoms (Å ²)	73.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.94% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

5 Model quality ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: BMA, NAG

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z > 5$	RMSZ	$\# Z > 5$
1	A	0.51	1/2591 (0.0%)	0.68	0/3517
1	B	0.49	0/2562	0.69	0/3478
1	C	0.53	0/2599	0.74	1/3528 (0.0%)
1	D	0.44	0/2599	0.64	1/3528 (0.0%)
1	E	0.48	1/2591 (0.0%)	0.65	0/3517
1	F	0.44	1/2599 (0.0%)	0.62	0/3528
1	G	0.42	1/2599 (0.0%)	0.63	0/3528
1	H	0.50	1/2608 (0.0%)	0.69	0/3540
1	I	0.45	1/2599 (0.0%)	0.65	0/3528
2	a	0.44	0/1415	0.63	0/1900
2	b	0.45	0/1369	0.64	0/1837
2	c	0.44	0/1424	0.60	0/1912
2	d	0.43	0/1424	0.58	0/1912
2	e	0.48	0/1424	0.62	0/1912
2	f	0.47	0/1424	0.62	0/1912
2	g	0.45	0/1424	0.62	0/1912
2	h	0.55	0/1381	0.65	0/1852
2	i	0.56	0/1424	0.67	0/1912
3	J	0.49	0/893	0.69	0/1213
3	L	0.60	0/1618	0.78	1/2209 (0.0%)
3	N	0.63	0/1689	0.83	1/2304 (0.0%)
3	P	0.42	0/1237	0.59	0/1682
3	R	0.45	0/1629	0.66	0/2224
3	T	0.39	0/1620	0.61	0/2212
3	V	0.44	0/905	0.60	0/1231
3	X	0.51	0/1261	0.70	0/1713
3	Z	0.42	0/1294	0.64	0/1760
4	K	0.43	0/843	0.72	1/1146 (0.1%)
4	M	0.57	0/1695	0.75	0/2302
4	O	0.67	0/1695	0.85	1/2302 (0.0%)
4	Q	0.44	0/1239	0.65	0/1682
4	S	0.45	0/1676	0.66	1/2277 (0.0%)

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
4	U	0.40	0/1676	0.60	0/2277
4	W	0.45	0/835	0.62	0/1135
4	Y	0.51	0/870	0.75	1/1181 (0.1%)
4	z	0.40	0/930	0.62	1/1261 (0.1%)
All	All	0.48	6/59661 (0.0%)	0.67	9/80864 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	G	0	1
2	b	0	2
2	f	0	1
2	i	0	2
3	J	0	2
3	L	0	2
3	N	0	3
3	R	0	1
3	X	0	1
3	Z	0	1
4	K	0	4
4	M	0	2
4	O	0	1
4	Q	0	2
4	S	0	1
4	U	0	1
4	W	0	1
4	Y	0	1
4	z	0	1
All	All	0	31

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	64	CYS	CB-SG	-7.01	1.70	1.82
1	E	64	CYS	CB-SG	-5.99	1.72	1.81
1	G	64	CYS	CB-SG	-5.53	1.72	1.81
1	F	64	CYS	CB-SG	-5.51	1.72	1.81
1	I	64	CYS	CB-SG	-5.20	1.73	1.81

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	220	ARG	NE-CZ-NH2	6.90	123.75	120.30
4	K	57	GLY	N-CA-C	-6.43	97.01	113.10
4	z	63	ARG	NE-CZ-NH1	-6.29	117.16	120.30
4	Y	63	ARG	NE-CZ-NH1	-5.84	117.38	120.30
3	L	102	PRO	N-CA-C	-5.32	98.28	112.10

There are no chirality outliers.

5 of 31 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	10	GLY	Peptide
1	G	9	PRO	Peptide
2	b	61	THR	Peptide
2	b	62	GLN	Peptide
2	f	60	ASN	Peptide

5.2 Close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2531	0	2485	45	0
1	B	2503	0	2463	56	0
1	C	2539	0	2497	57	0
1	D	2539	0	2499	48	0
1	E	2531	0	2487	56	0
1	F	2539	0	2496	48	0
1	G	2539	0	2499	57	0
1	H	2548	0	2502	63	0
1	I	2539	0	2499	56	0
2	a	1387	0	1293	0	0
2	b	1345	0	1254	0	0
2	c	1396	0	1299	0	0
2	d	1396	0	1299	0	0
2	e	1396	0	1299	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	f	1396	0	1299	0	0
2	g	1396	0	1299	0	0
2	h	1355	0	1263	0	0
2	i	1396	0	1299	0	0
3	J	871	0	832	34	0
3	L	1578	0	1539	42	0
3	N	1648	0	1610	39	0
3	P	1207	0	1157	41	0
3	R	1589	0	1549	54	0
3	T	1580	0	1540	52	0
3	V	883	0	848	25	0
3	X	1232	0	1192	41	0
3	Z	1262	0	1217	38	0
4	K	826	0	804	23	0
4	M	1661	0	1616	39	0
4	O	1661	0	1615	36	0
4	Q	1217	0	1180	30	0
4	S	1642	0	1602	49	0
4	U	1642	0	1602	52	0
4	W	818	0	793	27	0
4	Y	853	0	837	23	0
4	z	915	0	884	0	0
5	A	28	0	25	0	0
5	B	28	0	25	0	0
5	D	28	0	25	0	0
5	E	28	0	25	0	0
5	F	28	0	25	0	0
5	H	84	0	75	0	0
5	I	28	0	25	0	0
6	A	14	0	13	0	0
6	C	14	0	13	0	0
6	E	14	0	13	0	0
6	F	14	0	13	0	0
7	C	39	0	34	0	0
7	G	39	0	34	0	0
All	All	58742	0	56793	1046	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 11.

The worst 5 of 1046 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
3:N:171:GLN:HE21	3:N:177:SER:HB3	1.29	0.97
3:L:143:LYS:HA	3:L:177:SER:HB2	1.48	0.96
1:G:137:ARG:HG3	1:G:145:PRO:HG3	1.53	0.91
4:K:35:TRP:HB2	4:K:48:ILE:HG23	1.53	0.88
4:O:90:GLN:NE2	4:O:91:TYR:O	2.07	0.86

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	321/327 (98%)	292 (91%)	28 (9%)	1 (0%)	50	91
1	B	317/327 (97%)	283 (89%)	30 (10%)	4 (1%)	18	68
1	C	322/327 (98%)	288 (89%)	29 (9%)	5 (2%)	14	63
1	D	322/327 (98%)	293 (91%)	26 (8%)	3 (1%)	25	76
1	E	321/327 (98%)	291 (91%)	26 (8%)	4 (1%)	19	70
1	F	322/327 (98%)	288 (89%)	28 (9%)	6 (2%)	12	59
1	G	322/327 (98%)	285 (88%)	32 (10%)	5 (2%)	14	63
1	H	323/327 (99%)	286 (88%)	34 (10%)	3 (1%)	25	76
1	I	322/327 (98%)	287 (89%)	32 (10%)	3 (1%)	25	76
2	a	169/174 (97%)	143 (85%)	20 (12%)	6 (4%)	5	36
2	b	163/174 (94%)	135 (83%)	24 (15%)	4 (2%)	9	49
2	c	170/174 (98%)	147 (86%)	19 (11%)	4 (2%)	9	51
2	d	170/174 (98%)	147 (86%)	21 (12%)	2 (1%)	19	70
2	e	170/174 (98%)	154 (91%)	14 (8%)	2 (1%)	19	70
2	f	170/174 (98%)	144 (85%)	22 (13%)	4 (2%)	9	51
2	g	170/174 (98%)	148 (87%)	20 (12%)	2 (1%)	19	70
2	h	163/174 (94%)	143 (88%)	15 (9%)	5 (3%)	7	41
2	i	170/174 (98%)	144 (85%)	21 (12%)	5 (3%)	7	43

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	J	113/223 (51%)	80 (71%)	26 (23%)	7 (6%)	2	19
3	L	208/223 (93%)	186 (89%)	13 (6%)	9 (4%)	4	30
3	N	221/223 (99%)	200 (90%)	14 (6%)	7 (3%)	6	39
3	P	155/223 (70%)	134 (86%)	12 (8%)	9 (6%)	3	21
3	R	210/223 (94%)	185 (88%)	18 (9%)	7 (3%)	6	38
3	T	209/223 (94%)	184 (88%)	19 (9%)	6 (3%)	7	43
3	V	115/223 (52%)	106 (92%)	8 (7%)	1 (1%)	25	76
3	X	156/223 (70%)	129 (83%)	19 (12%)	8 (5%)	3	25
3	Z	159/223 (71%)	133 (84%)	16 (10%)	10 (6%)	2	18
4	K	104/214 (49%)	77 (74%)	17 (16%)	10 (10%)	1	7
4	M	212/214 (99%)	191 (90%)	18 (8%)	3 (1%)	16	66
4	O	212/214 (99%)	190 (90%)	18 (8%)	4 (2%)	12	59
4	Q	149/214 (70%)	125 (84%)	20 (13%)	4 (3%)	8	46
4	S	209/214 (98%)	182 (87%)	20 (10%)	7 (3%)	6	38
4	U	209/214 (98%)	177 (85%)	25 (12%)	7 (3%)	6	38
4	W	103/214 (48%)	89 (86%)	10 (10%)	4 (4%)	5	33
4	Y	107/214 (50%)	94 (88%)	9 (8%)	4 (4%)	5	34
4	z	113/214 (53%)	97 (86%)	11 (10%)	5 (4%)	4	29
All	All	7371/8442 (87%)	6457 (88%)	734 (10%)	180 (2%)	9	51

5 of 180 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	a	66	VAL
1	B	264	SER
2	c	61	THR
2	e	146	ASP
2	f	62	GLN

5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	285/289 (99%)	236 (83%)	49 (17%)	3	14
1	B	282/289 (98%)	228 (81%)	54 (19%)	2	11
1	C	286/289 (99%)	241 (84%)	45 (16%)	4	17
1	D	286/289 (99%)	239 (84%)	47 (16%)	3	15
1	E	285/289 (99%)	239 (84%)	46 (16%)	3	16
1	F	286/289 (99%)	235 (82%)	51 (18%)	2	13
1	G	286/289 (99%)	241 (84%)	45 (16%)	4	17
1	H	287/289 (99%)	239 (83%)	48 (17%)	3	14
1	I	286/289 (99%)	239 (84%)	47 (16%)	3	15
2	a	148/151 (98%)	128 (86%)	20 (14%)	6	26
2	b	144/151 (95%)	123 (85%)	21 (15%)	5	21
2	c	149/151 (99%)	137 (92%)	12 (8%)	17	56
2	d	149/151 (99%)	137 (92%)	12 (8%)	17	56
2	e	149/151 (99%)	134 (90%)	15 (10%)	11	41
2	f	149/151 (99%)	134 (90%)	15 (10%)	11	41
2	g	149/151 (99%)	136 (91%)	13 (9%)	15	51
2	h	145/151 (96%)	130 (90%)	15 (10%)	10	40
2	i	149/151 (99%)	135 (91%)	14 (9%)	13	45
3	J	91/185 (49%)	70 (77%)	21 (23%)	1	5
3	L	176/185 (95%)	142 (81%)	34 (19%)	2	10
3	N	185/185 (100%)	142 (77%)	43 (23%)	1	5
3	P	130/185 (70%)	104 (80%)	26 (20%)	2	9
3	R	177/185 (96%)	136 (77%)	41 (23%)	1	5
3	T	176/185 (95%)	138 (78%)	38 (22%)	1	7
3	V	93/185 (50%)	74 (80%)	19 (20%)	2	8
3	X	136/185 (74%)	106 (78%)	30 (22%)	1	7
3	Z	139/185 (75%)	109 (78%)	30 (22%)	1	7
4	K	93/190 (49%)	67 (72%)	26 (28%)	0	2
4	M	190/190 (100%)	158 (83%)	32 (17%)	3	14
4	O	190/190 (100%)	152 (80%)	38 (20%)	2	9
4	Q	140/190 (74%)	108 (77%)	32 (23%)	1	5
4	S	188/190 (99%)	151 (80%)	37 (20%)	2	10

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
4	U	188/190 (99%)	141 (75%)	47 (25%)	1	3
4	W	92/190 (48%)	73 (79%)	19 (21%)	2	8
4	Y	96/190 (50%)	76 (79%)	20 (21%)	2	8
4	z	105/190 (55%)	81 (77%)	24 (23%)	1	5
All	All	6485/7335 (88%)	5359 (83%)	1126 (17%)	3	13

5 of 1126 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	I	18	HIS
3	L	117	LYS
3	X	115	SER
1	I	146	SER
3	J	5	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 38 such sidechains are listed below:

Mol	Chain	Res	Type
1	H	295	HIS
2	i	125	GLN
4	Y	90	GLN
2	h	125	GLN
4	K	34	ASN

5.3.3 RNA ⓘ

There are no RNA chains in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

24 carbohydrates are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and

the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
5	NAG	A	401	1,5	12,14,15	0.68	0	15,19,21	1.24	2 (13%)
5	NAG	A	402	5	12,14,15	0.71	1 (8%)	15,19,21	0.82	0
5	NAG	B	401	1,5	12,14,15	0.64	0	15,19,21	1.10	2 (13%)
5	NAG	B	402	5	12,14,15	0.61	0	15,19,21	0.69	0
7	NAG	C	401	1,7	12,14,15	0.71	1 (8%)	15,19,21	1.58	3 (20%)
7	NAG	C	402	7	12,14,15	0.64	0	15,19,21	0.95	1 (6%)
7	BMA	C	403	7	10,11,12	0.76	0	11,15,17	2.22	2 (18%)
5	NAG	D	401	1,5	12,14,15	0.64	0	15,19,21	1.60	1 (6%)
5	NAG	D	402	5	12,14,15	0.71	1 (8%)	15,19,21	0.87	1 (6%)
5	NAG	E	402	1,5	12,14,15	0.76	0	15,19,21	1.43	3 (20%)
5	NAG	E	403	5	12,14,15	0.60	0	15,19,21	0.95	1 (6%)
5	NAG	F	401	1,5	12,14,15	0.81	0	15,19,21	1.67	3 (20%)
5	NAG	F	402	5	12,14,15	0.47	0	15,19,21	1.21	2 (13%)
7	NAG	G	401	1,7	12,14,15	0.69	0	15,19,21	0.96	1 (6%)
7	NAG	G	402	7	12,14,15	0.58	0	15,19,21	0.96	1 (6%)
7	BMA	G	403	7	10,11,12	0.77	0	11,15,17	0.71	0
5	NAG	H	401	1,5	12,14,15	0.60	0	15,19,21	1.39	1 (6%)
5	NAG	H	402	5	12,14,15	0.63	0	15,19,21	0.97	1 (6%)
5	NAG	H	403	1,5	12,14,15	0.62	0	15,19,21	1.62	3 (20%)
5	NAG	H	404	5	12,14,15	0.68	1 (8%)	15,19,21	0.78	0
5	NAG	H	405	1,5	12,14,15	0.59	0	15,19,21	1.63	1 (6%)
5	NAG	H	406	5	12,14,15	0.64	0	15,19,21	1.04	1 (6%)
5	NAG	I	401	1,5	12,14,15	0.57	0	15,19,21	0.95	1 (6%)
5	NAG	I	402	5	12,14,15	0.59	0	15,19,21	1.49	1 (6%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	NAG	A	401	1,5	-	0/6/23/26	0/1/1/1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	NAG	A	402	5	-	0/6/23/26	0/1/1/1
5	NAG	B	401	1,5	-	0/6/23/26	0/1/1/1
5	NAG	B	402	5	-	0/6/23/26	0/1/1/1
7	NAG	C	401	1,7	-	0/6/23/26	0/1/1/1
7	NAG	C	402	7	-	0/6/23/26	0/1/1/1
7	BMA	C	403	7	-	0/2/19/22	0/1/1/1
5	NAG	D	401	1,5	-	0/6/23/26	0/1/1/1
5	NAG	D	402	5	-	0/6/23/26	0/1/1/1
5	NAG	E	402	1,5	-	0/6/23/26	0/1/1/1
5	NAG	E	403	5	-	0/6/23/26	0/1/1/1
5	NAG	F	401	1,5	-	0/6/23/26	0/1/1/1
5	NAG	F	402	5	-	0/6/23/26	0/1/1/1
7	NAG	G	401	1,7	-	2/6/23/26	0/1/1/1
7	NAG	G	402	7	-	0/6/23/26	0/1/1/1
7	BMA	G	403	7	-	0/2/19/22	0/1/1/1
5	NAG	H	401	1,5	-	0/6/23/26	0/1/1/1
5	NAG	H	402	5	-	0/6/23/26	0/1/1/1
5	NAG	H	403	1,5	-	0/6/23/26	0/1/1/1
5	NAG	H	404	5	-	0/6/23/26	0/1/1/1
5	NAG	H	405	1,5	-	0/6/23/26	0/1/1/1
5	NAG	H	406	5	-	0/6/23/26	0/1/1/1
5	NAG	I	401	1,5	-	0/6/23/26	0/1/1/1
5	NAG	I	402	5	-	0/6/23/26	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	402	NAG	O5-C5	-2.21	1.41	1.45
5	D	402	NAG	O5-C5	-2.06	1.41	1.45
5	H	404	NAG	O5-C5	-2.03	1.41	1.45
7	C	401	NAG	O5-C5	-2.02	1.41	1.45

The worst 5 of 32 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	C	403	BMA	O5-C5-C6	6.74	114.05	106.98
5	D	401	NAG	O5-C5-C6	5.22	112.46	106.98
5	H	405	NAG	O5-C5-C6	5.06	112.29	106.98
5	I	402	NAG	O5-C5-C6	4.75	111.97	106.98
5	H	401	NAG	O5-C5-C6	4.60	111.81	106.98

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
7	G	401	NAG	C8-C7-N2-C2
7	G	401	NAG	O7-C7-N2-C2

There are no ring outliers.

5.6 Ligand geometry

4 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
6	NAG	A	403	1	12,14,15	0.54	0	15,19,21	1.34	1 (6%)
6	NAG	C	404	1	12,14,15	1.31	2 (16%)	15,19,21	1.05	0
6	NAG	E	401	1	12,14,15	0.56	0	15,19,21	1.06	0
6	NAG	F	403	1	12,14,15	0.52	0	15,19,21	1.39	1 (6%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	NAG	A	403	1	-	0/6/23/26	0/1/1/1
6	NAG	C	404	1	-	0/6/23/26	0/1/1/1
6	NAG	E	401	1	-	0/6/23/26	0/1/1/1
6	NAG	F	403	1	-	0/6/23/26	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	C	404	NAG	O5-C5	2.97	1.50	1.45
6	C	404	NAG	C2-N2	-2.35	1.43	1.46

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	F	403	NAG	O5-C5-C6	4.29	111.48	106.98
6	A	403	NAG	O5-C5-C6	4.13	111.32	106.98

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	323/327 (98%)	-0.30	0 100 100	20, 53, 91, 130	0
1	B	319/327 (97%)	-0.20	0 100 100	25, 60, 101, 142	0
1	C	324/327 (99%)	-0.22	0 100 100	21, 58, 98, 133	0
1	D	324/327 (99%)	-0.20	1 (0%) 91 58	38, 66, 99, 168	0
1	E	323/327 (98%)	-0.23	1 (0%) 91 58	45, 68, 97, 125	0
1	F	324/327 (99%)	-0.12	1 (0%) 91 58	45, 76, 107, 138	0
1	G	324/327 (99%)	0.00	0 100 100	46, 82, 124, 162	0
1	H	325/327 (99%)	-0.32	0 100 100	31, 53, 82, 122	0
1	I	324/327 (99%)	-0.15	0 100 100	42, 75, 107, 127	0
2	a	171/174 (98%)	-0.10	0 100 100	25, 75, 130, 168	0
2	b	167/174 (95%)	-0.05	0 100 100	24, 79, 138, 179	0
2	c	172/174 (98%)	-0.12	1 (0%) 86 41	30, 79, 134, 176	0
2	d	172/174 (98%)	0.05	1 (0%) 86 41	43, 91, 149, 196	0
2	e	172/174 (98%)	-0.11	0 100 100	38, 69, 123, 176	0
2	f	172/174 (98%)	-0.09	2 (1%) 75 26	46, 74, 125, 195	0
2	g	172/174 (98%)	-0.02	0 100 100	43, 81, 145, 169	0
2	h	167/174 (95%)	-0.26	0 100 100	35, 49, 85, 134	0
2	i	172/174 (98%)	-0.25	0 100 100	33, 55, 122, 187	0
3	J	115/223 (51%)	0.33	2 (1%) 67 19	46, 94, 132, 166	0
3	L	212/223 (95%)	-0.27	0 100 100	20, 38, 99, 134	0
3	N	223/223 (100%)	-0.31	0 100 100	14, 28, 73, 129	0
3	P	161/223 (72%)	0.12	2 (1%) 75 26	52, 90, 147, 172	0
3	R	214/223 (95%)	-0.02	0 100 100	45, 83, 114, 128	0
3	T	213/223 (95%)	0.32	10 (4%) 30 6	62, 107, 153, 233	0

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
3	V	117/223 (52%)	0.62	7 (5%) 21 5	66, 122, 174, 221	0
3	X	164/223 (73%)	-0.02	2 (1%) 75 26	35, 67, 110, 147	0
3	Z	167/223 (74%)	-0.04	1 (0%) 86 41	59, 95, 128, 180	0
4	K	106/214 (49%)	0.26	2 (1%) 64 18	63, 111, 145, 173	0
4	M	214/214 (100%)	-0.22	0 100 100	19, 42, 96, 150	0
4	O	214/214 (100%)	-0.38	0 100 100	16, 30, 52, 70	0
4	Q	157/214 (73%)	0.05	3 (1%) 64 18	49, 90, 134, 161	0
4	S	211/214 (98%)	-0.09	1 (0%) 88 46	52, 77, 123, 185	0
4	U	211/214 (98%)	0.18	1 (0%) 88 46	66, 104, 141, 162	0
4	W	105/214 (49%)	-0.06	0 100 100	57, 88, 125, 148	0
4	Y	109/214 (50%)	-0.31	0 100 100	36, 51, 76, 98	0
4	z	119/214 (55%)	0.17	0 100 100	61, 98, 140, 161	0
All	All	7479/8442 (88%)	-0.10	38 (0%) 88 46	14, 70, 129, 233	0

The worst 5 of 38 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	T	185	PRO	3.9
4	S	154	LEU	3.7
4	Q	134	CYS	3.6
1	D	9	PRO	3.5
3	J	90	TYR	3.5

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(\AA^2)	Q<0.9
7	NAG	C	402	14/15	0.30	-	88,117,138,139	0
5	NAG	I	401	14/15	0.23	-	68,81,87,88	0
5	NAG	H	406	14/15	0.49	-	148,157,164,167	0
7	BMA	G	403	11/12	0.17	-	117,121,132,137	0
7	NAG	G	402	14/15	0.19	-	93,112,127,132	0
5	NAG	H	404	14/15	0.32	-	94,108,128,135	0
5	NAG	E	402	14/15	0.18	-	60,86,97,102	0
5	NAG	H	402	14/15	0.44	-	98,107,122,125	0
5	NAG	F	402	14/15	0.26	-	97,113,122,124	0
5	NAG	A	402	14/15	0.22	-	90,107,119,121	0
7	NAG	G	401	14/15	0.17	-	84,97,101,102	0
7	NAG	C	401	14/15	0.26	-	56,81,94,108	0
5	NAG	H	405	14/15	0.25	-	88,112,125,131	0
7	BMA	C	403	11/12	0.30	-	109,140,145,148	0
5	NAG	D	402	14/15	0.16	-	85,111,119,120	0
5	NAG	D	401	14/15	0.09	-	61,84,89,97	0
5	NAG	B	402	14/15	0.22	-	125,133,136,140	0
5	NAG	I	402	14/15	0.23	-	82,103,118,122	0
5	NAG	H	401	14/15	0.15	-	50,75,92,93	0
5	NAG	F	401	14/15	0.24	-	60,82,88,96	0
5	NAG	A	401	14/15	0.22	-	72,88,97,98	0
5	NAG	B	401	14/15	0.15	-	63,92,107,114	0
5	NAG	H	403	14/15	0.17	-	61,80,90,101	0
5	NAG	E	403	14/15	0.24	-	101,113,121,122	0

6.4 Ligands

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(\AA^2)	Q<0.9
6	NAG	E	401	14/15	0.21	-	80,99,105,113	0
6	NAG	C	404	14/15	0.39	-	116,126,129,130	0
6	NAG	A	403	14/15	0.31	-	75,97,108,109	0
6	NAG	F	403	14/15	0.25	-	68,81,94,96	0

6.5 Other polymers ⓘ

There are no such residues in this entry.